

Let us do the twist again

Oskar Maria Baksalary · Götz Trenkler ·
Erkki Liski

Received: 8 June 2012 / Revised: 25 October 2012 / Published online: 17 March 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Krämer (Sankhyā 42:130–131, 1980) posed the following problem: “Which are the \mathbf{y} , given \mathbf{X} and \mathbf{V} , such that OLS and Gauss–Markov are equal?”. In other words, the problem aimed at identifying those vectors \mathbf{y} for which the ordinary least squares (OLS) and Gauss–Markov estimates of the parameter vector $\boldsymbol{\beta}$ coincide under the general Gauss–Markov model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. The problem was later called a “twist” to Kruskal’s Theorem, which provides conditions necessary and sufficient for the OLS and Gauss–Markov estimates of $\boldsymbol{\beta}$ to be equal. The present paper focuses on a similar problem to the one posed by Krämer in the aforementioned paper. However, instead of the estimation of $\boldsymbol{\beta}$, we consider the estimation of the systematic part $\mathbf{X}\boldsymbol{\beta}$, which is a natural consequence of relaxing the assumption that \mathbf{X} and \mathbf{V} are of full (column) rank made by Krämer. Further results, dealing with the Euclidean distance between the best linear unbiased estimator (BLUE) and the ordinary least squares estimator (OLSE) of $\mathbf{X}\boldsymbol{\beta}$, as well as with an equality between BLUE and OLSE are also provided. The calculations are mostly based on a joint partitioned representation of a pair of orthogonal projectors.

O. M. Baksalary (✉)

Faculty of Physics, Adam Mickiewicz University, ul., Umultowska 85, 61-614 Poznań, Poland
e-mail: obaksalary@gmail.com

G. Trenkler

Faculty of Statistics, Dortmund University of Technology, Vogelpothsweg 87, 44221 Dortmund, Germany
e-mail: trenkler@statistik.tu-dortmund.de

E. Liski

School of Information Sciences, University of Tampere, 33014 Tampere, Finland
e-mail: Erkki.Liski@uta.fi

Keywords Gauss–Markov model · Kruskal’s Theorem · Best linear unbiased estimator · Ordinary least squares unbiased estimator · Orthogonal projector

Mathematics Subject Classification (2000) 62J05 · 62H12 · 15A24 · 15A04

1 Introduction

Let us consider the general Gauss–Markov model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1)$$

where \mathbf{y} is an $n \times 1$ observable random vector, \mathbf{X} is a known $n \times p$ model matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and \mathbf{u} is an $n \times 1$ random error vector. The expectation vector and the covariance matrix of \mathbf{u} are $\mathbf{E}(\mathbf{u}) = \mathbf{0}$ and $\text{Cov}(\mathbf{u}) = \sigma^2 \mathbf{V}$, respectively, where $\sigma^2 > 0$ is an unknown constant and \mathbf{V} is a known $n \times n$ nonnegative definite matrix. Both \mathbf{X} and \mathbf{V} may be rank deficient. It is assumed beforehand that the model (1) is consistent, i.e., $\mathbf{y} \in \mathcal{R}(\mathbf{X} : \mathbf{V})$, where $\mathcal{R}(\cdot)$ stands for the column space of a matrix argument and $(\mathbf{X} : \mathbf{V})$ denotes the $n \times (p + n)$ columnwise partitioned matrix obtained by juxtaposing matrices \mathbf{X} and \mathbf{V} ; cf. Rao (1973, p. 297) or Puntanen et al. (2011, pp. 43, 125).

In his paper, Krämer (1980, p. 130) posed the following problem: “Which are the \mathbf{y} , given \mathbf{X} and \mathbf{V} , such that ordinary least squares (OLS) and Gauss–Markov are equal?” In other words, the problem aimed at identifying those vectors \mathbf{y} for which the OLS and Gauss–Markov estimates of the parameter vector $\boldsymbol{\beta}$ coincide. Referring to this problem, in a follow-up paper Krämer et al. (1996) called this a “twist” to Kruskal’s Theorem (Kruskal 1968), which provides conditions necessary and sufficient for the OLS and Gauss–Markov estimates of $\boldsymbol{\beta}$ to be equal. In Krämer et al. (1996) “another twist” to Kruskal’s Theorem is dealt with, and rather than asking when is the OLS equal to the Gauss–Markov for the full regression vector $\boldsymbol{\beta}$, a condition for the equality of the OLS and Gauss–Markov for a subparameter of $\boldsymbol{\beta}$ is provided. A more general “final twist” is considered in Jaeger and Krämer (1998), where the single vectors \mathbf{y} are characterized that yield identical OLS and Gauss–Markov estimators for such a subparameter.

Inspired by Jaeger and Krämer (1998), Krämer (1980), and Krämer et al. (1996), in what follows we “do the twist again”. However, unlike in the three papers, we do not assume that \mathbf{X} and \mathbf{V} are of full (column) rank, which means that the vector $\boldsymbol{\beta}$ is not necessarily unbiasedly estimable. For this reason, instead of the estimation of $\boldsymbol{\beta}$, we consider the estimation of the systematic part $\mathbf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. Note that this parameter function always has a linear unbiased estimator, namely \mathbf{y} itself.

An important role in the subsequent considerations will be played by the notion of a projector. It is known that any $n \times n$ idempotent matrix, say $\mathbf{F} \in \mathbb{R}^{n \times n}$, is an oblique projector onto its column space $\mathcal{R}(\mathbf{F})$ along its null space $\mathcal{N}(\mathbf{F})$, where $\mathcal{R}(\mathbf{F}) \oplus \mathcal{N}(\mathbf{F}) = \mathbb{R}^n$.¹ Among many conditions characterizing idempotent matrices one finds for instance: $\mathbf{F}^2 = \mathbf{F} \Leftrightarrow \mathcal{R}(\mathbf{F}) = \mathcal{N}(\bar{\mathbf{F}}) \Leftrightarrow \mathcal{R}(\bar{\mathbf{F}}) = \mathcal{N}(\mathbf{F})$, where $\bar{\mathbf{F}} =$

$\mathbf{I}_n - \mathbf{F}$. When idempotent \mathbf{F} projects onto $\mathcal{R}(\mathbf{F})$ along the orthogonal complement of $\mathcal{R}(\mathbf{F})$, then it is called an orthogonal projector. It can be verified that \mathbf{F} is an orthogonal projector if and only if it is both idempotent and symmetric, i.e., $\mathbf{F}^2 = \mathbf{F} = \mathbf{F}'$. Projectors are widely used in Statistics and Econometrics as a basic tool for estimation and test procedures.

Let \mathbf{G} be an $n \times n$ matrix. An estimator $\mathbf{G}\mathbf{y}$ for $\mathbf{X}\boldsymbol{\beta}$, which is unbiased and of minimal covariance matrix in the Löwner sense fulfills the conditions

$$\mathbf{G}\mathbf{H} = \mathbf{H} \quad \text{and} \quad \mathbf{G}\mathbf{V}\mathbf{M} = \mathbf{0}, \quad (2)$$

where $\mathbf{H} = \mathbf{X}\mathbf{X}^\dagger$ and $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$ are the orthogonal projectors onto, respectively, $\mathcal{R}(\mathbf{X})$, the column space of \mathbf{X} , and the orthogonal complement of $\mathcal{R}(\mathbf{X})$ which coincides with $\mathcal{N}(\mathbf{X}')$, the null space of \mathbf{X}' . The symbol \mathbf{X}^\dagger denotes the Moore–Penrose inverse of \mathbf{X} . The conditions (2) can be rewritten as

$$\mathbf{G}(\mathbf{H} : \mathbf{P}_{\mathbf{V}\mathbf{M}}) = (\mathbf{H} : \mathbf{0}), \quad (3)$$

where $\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{V}\mathbf{M}(\mathbf{V}\mathbf{M})^\dagger$ is the orthogonal projector onto $\mathcal{R}(\mathbf{V}\mathbf{M})$. It was pointed out in Baksalary and Trenkler (2012, Remark 3.1) that Eq. (3) always has a solution \mathbf{G} and that each \mathbf{G} satisfying (3) yields a representation of the best linear unbiased estimator $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ of $\mathbf{X}\boldsymbol{\beta}$. All these representations coincide; see Groß (2004, Corollary 3). In the Appendix given below it is demonstrated that there may exist, however, quite useless versions of $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$. To avoid this discrepancy subsequently we strengthen the consistency condition $\mathbf{y} \in \mathcal{R}(\mathbf{X} : \mathbf{V})$ to

$$\mathcal{R}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^{n,1}. \quad (4)$$

Then, according to Groß (2004, Corollary 4), $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ is unique.

In the next section some representations of the best linear unbiased estimator (BLUE) and the ordinary least squares estimator (OLSE) are provided, whereas Sect. 3 deals with “another twist” to Kruskal’s Theorem, which was briefly mentioned above. Section 4 is concerned with bounds for the Euclidean distance between BLUE and OLSE of $\mathbf{X}\boldsymbol{\beta}$, and the last section of the paper revisits the problem of when BLUE equals OLSE.

2 Representations of BLUE and OLSE

Let \mathbf{P} be an orthogonal projector in $\mathbb{R}^{n,1}$, i.e., an $n \times n$ real symmetric idempotent matrix. Assume that the rank of \mathbf{P} is r . It is known that there exists an orthogonal matrix \mathbf{U} such that

$$\mathbf{P} = \mathbf{U} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}'; \quad (5)$$

see Trenkler (1994, Theorem 13). Any other orthogonal projector of the same size, say $\mathbf{Q} \in \mathbb{R}^{n \times n}$, can be represented as

$$\mathbf{Q} = \mathbf{U} \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{pmatrix} \mathbf{U}', \quad (6)$$

with symmetric matrices \mathbf{A} and \mathbf{D} of orders r and $n - r$, respectively.

In what follows we assume that $\mathbf{H} = \mathbf{X}\mathbf{X}^\dagger$ is represented by \mathbf{P} of the form (5) and $\mathbf{P}_{\mathbf{VM}}$ is represented by \mathbf{Q} defined in (6), i.e., $\mathbf{P} = \mathbf{H} = \mathbf{X}\mathbf{X}^\dagger$ and $\mathbf{Q} = \mathbf{P}_{\mathbf{VM}} = \mathbf{VM}(\mathbf{VM})^\dagger$. It can be verified that $\mathbf{T} = (\bar{\mathbf{P}}_{\mathbf{VM}}\mathbf{H})^\dagger$, where $\bar{\mathbf{P}}_{\mathbf{VM}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{VM}}$, is an idempotent matrix; see Greville (1974, p. 830). From (5) and (6) we obtain

$$\mathbf{T} = \mathbf{U} \begin{pmatrix} \mathbf{P}_{\bar{\mathbf{A}}} & -\mathbf{BD}^\dagger \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}',$$

where $\mathbf{P}_{\bar{\mathbf{A}}}$ is the orthogonal projector onto the column space of $\bar{\mathbf{A}} = \mathbf{I}_r - \mathbf{A}$. It follows that \mathbf{T} is the oblique projector onto $\mathcal{R}(\mathbf{H}) \cap [\mathcal{N}(\mathbf{H}) + \mathcal{N}(\mathbf{P}_{\mathbf{VM}})]$ along $\mathcal{R}(\mathbf{P}_{\mathbf{VM}}) \overset{\perp}{\oplus} [\mathcal{N}(\mathbf{H}) \cap \mathcal{N}(\mathbf{P}_{\mathbf{VM}})]$, where $\overset{\perp}{\oplus}$ indicates that the two subspaces involved in the direct sum are orthogonal; see Baksalary and Trenkler (2010, Theorem 2). From

$$\mathcal{R}(\mathbf{H}) \cap \mathcal{R}(\mathbf{VM}) = \mathcal{R}(\mathbf{H}) \cap \mathcal{R}(\mathbf{P}_{\mathbf{VM}}) = \{\mathbf{0}\} \quad (7)$$

(see Baksalary and Trenkler 2009, Theorem 1), we arrive at $\mathcal{N}(\mathbf{H}) + \mathcal{N}(\mathbf{P}_{\mathbf{VM}}) = \mathbb{R}^{n,1}$, which leads to the conclusion that \mathbf{T} is the oblique projector onto $\mathcal{R}(\mathbf{H})$ along $\mathcal{R}(\mathbf{P}_{\mathbf{VM}}) \overset{\perp}{\oplus} [\mathcal{N}(\mathbf{H}) \cap \mathcal{N}(\mathbf{P}_{\mathbf{VM}})]$. Furthermore, it follows that \mathbf{T} takes the form

$$\mathbf{T} = \mathbf{U} \begin{pmatrix} \mathbf{I}_r & -\mathbf{BD}^\dagger \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}'; \quad (8)$$

see Baksalary and Trenkler (2010, Sect. 2).

It is well known that (7) ensures that Eq. (3) is solvable. One of the solutions, namely $\mathbf{T} = (\bar{\mathbf{P}}_{\mathbf{VM}}\mathbf{H})^\dagger$, gives a representation of the BLUE for $\mathbf{X}\boldsymbol{\beta}$, i.e., $\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{T}\mathbf{y}$. There is a number of further expressions for the BLUE (see Baksalary and Trenkler 2009, Sect. 4), but they all coincide by the assumption (4). Observe that the OLSE of $\mathbf{X}\boldsymbol{\beta}$ is $\text{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{H}\mathbf{y}$.

3 Another twist

As in Krämer (1980), we consider the problem of identifying those observation vectors \mathbf{y} which yield the same value of $\text{OLSE}(\mathbf{X}\boldsymbol{\beta})$ and $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$. This amounts to an analysis of the subspace \mathcal{L} of $\mathbb{R}^{n,1}$ which is the null space of $\mathbf{H} - \mathbf{T}$, i.e., $\mathcal{L} = \mathcal{N}(\mathbf{H} - \mathbf{T})$. For this purpose the following result is useful.

Lemma 1 *Let \mathbf{R} and \mathbf{S} be idempotent matrices of the same size. Then:*

- (i) $\mathcal{N}(\mathbf{S} - \mathbf{R}) = \mathcal{N}(\mathbf{S}\bar{\mathbf{R}}) \cap \mathcal{N}(\bar{\mathbf{S}}\mathbf{R})$,
- (ii) $\mathcal{N}(\mathbf{RS}) = \mathcal{N}(\mathbf{S}) \oplus [\mathcal{N}(\mathbf{R}) \cap \mathcal{R}(\mathbf{S})]$.

Proof For a proof see [Baksalary and Trenkler \(2013, Theorems 1 and 9\)](#). \square

Lemma 1 leads to the following result.

Theorem 1 *Under the model (1), let $\mathbf{H} = \mathbf{X}\mathbf{X}^\dagger$, $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$, and $\mathbf{T} = (\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})^\dagger$. Then*

$$\mathcal{L} = \mathcal{N}(\mathbf{H} - \mathbf{T}) = \mathcal{N}(\mathbf{T}\mathbf{M}).$$

Proof Lemma 1 yields

$$\mathcal{N}(\mathbf{H} - \mathbf{T}) = \mathcal{N}(\mathbf{T}\bar{\mathbf{H}}) \cap \mathcal{N}(\bar{\mathbf{T}}\mathbf{H}) = \mathcal{N}(\mathbf{T}\mathbf{M}) \cap \mathcal{N}(\bar{\mathbf{T}}\mathbf{H}).$$

Another relevant fact is that with \mathbf{H} of the form (5) and \mathbf{T} given in (8) we directly get $\bar{\mathbf{T}}\mathbf{H} = \mathbf{0}$. \square

The vectors belonging to the subspace $\mathcal{L} = \mathcal{N}(\mathbf{H} - \mathbf{T})$ can be explicitly written as

$$\mathcal{L} = \{[\mathbf{I}_n - (\mathbf{T}\mathbf{M})^\dagger\mathbf{T}\mathbf{M}]\mathbf{z} : \mathbf{z} \in \mathbb{R}^{n,1}\},$$

as the solutions to the equation $\mathbf{T}\mathbf{M}\mathbf{z} = \mathbf{0}$. Observe also that $\mathcal{N}(\mathbf{T}\mathbf{M}) \supseteq \mathcal{N}(\mathbf{M}) = \mathcal{R}(\mathbf{H})$. This means that, inter alia, all vectors belonging to $\mathcal{R}(\mathbf{H}) = \mathcal{R}(\mathbf{X})$ result in estimates such that $\text{BLUE}(\mathbf{X}\beta) = \text{OLSE}(\mathbf{X}\beta)$, for example $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$. This does not come as a surprise, since by (5) and (8) it follows that $\mathbf{T}\mathbf{H} = \mathbf{H}$.

Further characterization of the subspace $\mathcal{L} = \mathcal{N}(\mathbf{H} - \mathbf{T})$ is established in the theorem below.

Theorem 2 *Under the model (1), let $\mathbf{H} = \mathbf{X}\mathbf{X}^\dagger$, $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$, and $\mathbf{T} = (\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})^\dagger$. Then*

$$\mathcal{L} = \mathcal{R}(\mathbf{H}) \overset{\perp}{\oplus} [\mathcal{N}(\mathbf{H}\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}) \cap \mathcal{N}(\mathbf{H})].$$

Proof By Theorem 1 we have $\mathcal{L} = \mathcal{N}(\mathbf{T}\mathbf{M})$. Hence, Lemma 1 implies

$$\mathcal{N}(\mathbf{T}\mathbf{M}) = \mathcal{N}(\mathbf{M}) \oplus [\mathcal{N}(\mathbf{T}) \cap \mathcal{R}(\mathbf{M})] = \mathcal{R}(\mathbf{H}) \oplus [\mathcal{N}(\mathbf{T}) \cap \mathcal{N}(\mathbf{H})].$$

Now

$$\mathcal{N}(\mathbf{T}) = \mathcal{N}[(\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})^\dagger] = \mathcal{N}[(\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})'] = \mathcal{N}(\mathbf{H}\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}),$$

which completes the proof. \square

The result of Theorem 2 looks somehow different than that of Groß et al. (2001, Theorem 9), for setting there $\mathbf{C} = \mathbf{I}_n$ gives

$$\mathcal{L} = \mathcal{R}(\mathbf{H}) \oplus [\mathcal{R}(\mathbf{VM}) \cap \mathcal{N}(\mathbf{H})]. \quad (9)$$

This discrepancy can be explained on account of the identity

$$\mathcal{N}(\mathbf{H}\bar{\mathbf{P}}_{\mathbf{VM}}) = \mathcal{R}(\mathbf{P}_{\mathbf{VM}}) \overset{\perp}{\oplus} [\mathcal{N}(\mathbf{H}) \cap \mathcal{N}(\mathbf{P}_{\mathbf{VM}})], \quad (10)$$

following from Lemma 1. The subspace of Theorem 2 coincides with (9) when $\mathcal{N}(\mathbf{H}) \cap \mathcal{N}(\mathbf{P}_{\mathbf{VM}}) = \{\mathbf{0}\}$, which is equivalent to $\mathcal{R}(\mathbf{H}) + \mathcal{R}(\mathbf{P}_{\mathbf{VM}}) = \mathbb{R}^{n,1}$ or $\mathcal{R}(\mathbf{H} : \mathbf{VM}) = \mathcal{R}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^{n,1}$; see Puntanen et al. (2011, Proposition 5.1). However, the latter condition, given above as (4), was assumed to be valid in the whole paper. Thus, we may state what follows.

Corollary 1 *Under the model (1), let $\mathbf{H} = \mathbf{XX}^\dagger$ and $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. Then*

$$\mathcal{L} = \mathcal{R}(\mathbf{H}) \overset{\perp}{\oplus} [\mathcal{R}(\mathbf{VM}) \cap \mathcal{N}(\mathbf{H})].$$

Corollary 1 corresponds to Krämer's (1980, Theorem), where the identity $\text{BLUE}(\boldsymbol{\beta}) = \text{OLSE}(\boldsymbol{\beta})$ is explored under the assumption that \mathbf{X} and \mathbf{V} are of full (column) rank.

Recall that the projector \mathbf{P} introduced in (5) was determined by the model matrix \mathbf{X} , for $\mathbf{P} = \mathbf{H} = \mathbf{XX}^\dagger$. In consequence, rank of \mathbf{P} coincides with the ranks of \mathbf{H} and \mathbf{X} , i.e., $r = \text{rank}(\mathbf{H}) = \text{rank}(\mathbf{X})$. Consider now an oblique projector of rank r having the form

$$\mathbf{L} = \mathbf{U} \begin{pmatrix} \mathbf{I}_r & \mathbf{K} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}', \quad (11)$$

with $\mathbf{K} \in \mathbb{R}^{r \times n-r}$. It was shown by Baksalary and Trenkler (2011, Sect. 3) that when $\mathbf{K} = -\mathbf{W}_{12}(\mathbf{DW}_{22}\mathbf{D})^\dagger$, where $\mathbf{D} \in \mathbb{R}^{n-r \times n-r}$ is a symmetric idempotent matrix and $\mathbf{W}_{12} \in \mathbb{R}^{r \times n-r}$ and $\mathbf{W}_{22} \in \mathbb{R}^{n-r \times n-r}$ originate from the representation of \mathbf{V} given by

$$\mathbf{V} = \mathbf{U} \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}'_{12} & \mathbf{W}_{22} \end{pmatrix} \mathbf{U}',$$

then \mathbf{Ly} is an unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$ whose efficiency lies between that of $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and $\text{OLSE}(\mathbf{X}\boldsymbol{\beta})$. In what follows we identify those observation vectors \mathbf{y} which yield the same estimators, compared to $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and $\text{OLSE}(\mathbf{X}\boldsymbol{\beta})$. The resulting formulas give an impression how close the three estimators can be.

Theorem 3 *Under the model (1), let $\mathbf{H} = \mathbf{XX}^\dagger$, $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$, and $\mathbf{T} = (\bar{\mathbf{P}}_{\mathbf{VM}}\mathbf{H})^\dagger$. Moreover, let \mathbf{L} be of the form (11). Then:*

- (i) $\mathcal{N}(\mathbf{H} - \mathbf{L}) = \mathcal{N}(\mathbf{LM}) = \mathcal{R}(\mathbf{L}) \oplus [\mathcal{N}(\mathbf{H}) \cap \mathcal{N}(\mathbf{L})],$
- (ii) $\mathcal{N}(\mathbf{T} - \mathbf{L}) = \mathcal{N}(\mathbf{T}\bar{\mathbf{L}}) = \mathcal{R}(\mathbf{T}) \oplus [\mathcal{N}(\mathbf{L}) \cap \mathcal{N}(\mathbf{T})].$

Proof From Lemma 1 we have $\mathcal{N}(\mathbf{H} - \mathbf{L}) = \mathcal{N}(\mathbf{H}\bar{\mathbf{L}}) \cap \mathcal{N}(\bar{\mathbf{H}}\mathbf{L})$. Representations (5) and (11) yield $\bar{\mathbf{H}}\mathbf{L} = \mathbf{0}$, whence, again by Lemma 1,

$$\mathcal{N}(\mathbf{H} - \mathbf{L}) = \mathcal{N}(\mathbf{H}\bar{\mathbf{L}}) = \mathcal{N}(\bar{\mathbf{L}}) \oplus [\mathcal{N}(\mathbf{H}) \cap \mathcal{R}(\bar{\mathbf{L}})] = \mathcal{R}(\mathbf{L}) \oplus [\mathcal{N}(\mathbf{H}) \cap \mathcal{N}(\mathbf{L})].$$

Since Theorem 1 ensures that $\mathcal{N}(\mathbf{H} - \mathbf{L}) = \mathcal{N}(\mathbf{L}\bar{\mathbf{H}}) = \mathcal{N}(\mathbf{L}\mathbf{M})$, point (i) of the theorem is established.

To derive the second part of the theorem, note that Lemma 1 entails $\mathcal{N}(\mathbf{T} - \mathbf{L}) = \mathcal{N}(\mathbf{L}\bar{\mathbf{T}}) \cap \mathcal{N}(\bar{\mathbf{L}}\mathbf{T})$. Similarly as in the proof of point (i), we obtain $\bar{\mathbf{L}}\mathbf{T} = \mathbf{0}$ which implies $\mathcal{N}(\mathbf{T} - \mathbf{L}) = \mathcal{N}(\mathbf{L}\bar{\mathbf{T}})$, and thus, by Lemma 1,

$$\mathcal{N}(\mathbf{T} - \mathbf{L}) = \mathcal{N}(\mathbf{L}\bar{\mathbf{T}}) = \mathcal{N}(\bar{\mathbf{T}}) \oplus [\mathcal{N}(\mathbf{L}) \cap \mathcal{R}(\bar{\mathbf{T}})] = \mathcal{R}(\mathbf{T}) \oplus [\mathcal{N}(\mathbf{L}) \cap \mathcal{N}(\mathbf{T})].$$

On the other hand, from Theorem 1 we have $\mathcal{N}(\mathbf{T} - \mathbf{L}) = \mathcal{N}(\mathbf{T}\bar{\mathbf{L}})$, which completes the proof. \square

4 Bounds for the Euclidean distance

Baksalary and Kala (1980) derived a bound for $\|\mu^* - \hat{\mu}\|$, where $\mu^* = \text{OLSE}(\mathbf{X}\beta)$ and $\hat{\mu} = \text{BLUE}(\mathbf{X}\beta)$, and $\|\cdot\|$ denotes the Euclidean norm. To be precise (Baksalary and Kala 1980, Theorem) reads

$$\|\mu^* - \hat{\mu}\| \leq (\gamma^{1/2}/\delta)\|\mathbf{y} - \mu^*\|,$$

where γ is the largest eigenvalue of $\mathbf{H}\mathbf{V}\mathbf{M}\mathbf{V}\mathbf{H}$ and δ is the smallest nonzero eigenvalue of $\mathbf{M}\mathbf{V}\mathbf{M}$. Subsequently, we provide an alternative bound, derived from the representation of $\text{BLUE}(\mathbf{X}\beta)$, using the oblique projector $\mathbf{T} = (\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})^\dagger$ given in (8).

Theorem 4 Let $\hat{\mu} = (\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})^\dagger \mathbf{y}$ with $(\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})^\dagger$ of the form (8) be one of the representations of $\text{BLUE}(\mathbf{X}\beta)$. If $\mu^* = \mathbf{H}\mathbf{y} = \text{OLSE}(\mathbf{X}\beta)$, then

$$\|\mu^* - \hat{\mu}\| \leq \tau_1(\mathbf{B}\mathbf{D}^\dagger)\|\mathbf{y}\|,$$

where $\tau_1(\mathbf{B}\mathbf{D}^\dagger)$ is the largest singular value of $\mathbf{B}\mathbf{D}^\dagger$.

Proof From

$$\mu^* = \mathbf{U} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}'\mathbf{y} \quad \text{and} \quad \hat{\mu} = \mathbf{U} \begin{pmatrix} \mathbf{I}_r & -\mathbf{B}\mathbf{D}^\dagger \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}'\mathbf{y}$$

we obtain

$$\mu^* - \hat{\mu} = \mathbf{U} \begin{pmatrix} \mathbf{0} & \mathbf{B}\mathbf{D}^\dagger \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}'\mathbf{y}.$$

Hence,

$$\|\mu^* - \hat{\mu}\|^2 = \mathbf{y}'\mathbf{U} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^\dagger \mathbf{B}' \mathbf{B} \mathbf{D}^\dagger \end{pmatrix} \mathbf{U}' \mathbf{y}.$$

In consequence, $\|\mu^* - \hat{\mu}\|^2 \leq \lambda_1(\mathbf{D}^\dagger \mathbf{B}' \mathbf{B} \mathbf{D}^\dagger) \|\mathbf{y}\|^2$, where $\lambda_1(\mathbf{D}^\dagger \mathbf{B}' \mathbf{B} \mathbf{D}^\dagger)$ is the largest eigenvalue of $\mathbf{D}^\dagger \mathbf{B}' \mathbf{B} \mathbf{D}^\dagger$. The assertion follows by taking square roots. \square

It is seen from Theorem 4 that $\mu^* = \hat{\mu}$ for all \mathbf{y} if and only if $\mathbf{B} \mathbf{D}^\dagger = \mathbf{0}$, or, equivalently, $\mathbf{B} = \mathbf{0}$, which means that \mathbf{H} and $\mathbf{P}_{\mathbf{V}\mathbf{M}}$ commute.

5 Equality of BLUE and OLSE

The commutativity $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H}$, just mentioned in the preceding section, is not contained in the standard catalogue of conditions necessary and sufficient for the equality $\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \text{OLSE}(\mathbf{X}\boldsymbol{\beta})$. Among the most important conditions equivalent to the equality are:

- (i) $\mathbf{H}\mathbf{V} = \mathbf{V}\mathbf{H}$,
- (ii) $\mathbf{H}\mathbf{V} = \mathbf{H}\mathbf{V}\mathbf{H}$,
- (iii) $\mathcal{R}(\mathbf{V}\mathbf{X}) = \mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\mathbf{V})$,
- (iv) $\mathbf{H}\mathbf{V}\mathbf{M} = \mathbf{0}$,
- (v) $\mathcal{R}(\mathbf{V}\mathbf{X}) \subseteq \mathcal{R}(\mathbf{X})$.

Note that the condition (v) can be rewritten as $\mathcal{R}(\mathbf{V}\mathbf{X}) = \mathcal{R}(\mathbf{X})$ when \mathbf{V} is nonsingular; see Krämer (1980) for a discussion related to Kruskal's theorem and Puntanen et al. (2011, Proposition 10.1). Motivated by Theorem 4 we obtain the following result.

Theorem 5 *Under the model (1), let $\mathbf{H} = \mathbf{X}\mathbf{X}^\dagger$ and $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. Then, the following conditions are equivalent:*

- (i) $\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \text{OLSE}(\mathbf{X}\boldsymbol{\beta})$,
- (ii) $\mathbf{H}\mathbf{V}\mathbf{M} = \mathbf{0}$,
- (iii) $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{0}$,
- (iv) $\mathbf{H} + \mathbf{P}_{\mathbf{V}\mathbf{M}}$ is an orthogonal projector,
- (v) $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H}$,
- (vi) $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}$ is an orthogonal projector.

Proof For the proof of the equivalence (i) \Leftrightarrow (ii) see Puntanen et al. (2011, Proposition 10.1).

To show that (ii) implies (iii) postmultiply $\mathbf{H}\mathbf{V}\mathbf{M} = \mathbf{0}$ by $(\mathbf{V}\mathbf{M})^\dagger$ and refer to $\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{V}\mathbf{M}(\mathbf{V}\mathbf{M})^\dagger$. To establish the reverse implication, note that the condition $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{0}$ entails $\mathbf{H}\mathbf{V}\mathbf{M}(\mathbf{V}\mathbf{M})^\dagger = \mathbf{0}$. Postmultiplying this equality by $\mathbf{V}\mathbf{M}$ leads to $\mathbf{H}\mathbf{V}\mathbf{M} = \mathbf{0}$.

The equivalence (iii) \Leftrightarrow (iv) is well known; see e.g., Rao and Mitra (1971, Theorem 5.1.2).

The fact that (iii) \Rightarrow (v) is visibly seen by taking the transpose of $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{0}$. For the proof of the reverse implication, recall that $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ exists if and only if Eq. (7)

are satisfied. By condition (v) we have $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{x} = \mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H}\mathbf{x}$ for any vector $\mathbf{x} \in \mathbb{R}^{n,1}$. Thus, $\mathbf{x} \in \mathcal{R}(\mathbf{H}) \cap \mathcal{R}(\mathbf{P}_{\mathbf{V}\mathbf{M}})$, whence $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{x} = \mathbf{0}$ for any \mathbf{x} , i.e., $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{0}$.

The part (v) \Leftrightarrow (vi) is also known in the literature; see e.g., Baksalary et al. (2002, Theorem). \square

Krämer (1980) showed how his theorem characterizing the vectors \mathbf{y} ensuring the coincidence of $\text{BLUE}(\boldsymbol{\beta})$ and $\text{OLSE}(\boldsymbol{\beta})$ can be used to prove Kruskal's Theorem. This is done in a similar fashion in the present set-up.

Theorem 6 *Under the model (1), let $\mathbf{H} = \mathbf{X}\mathbf{X}^\dagger$ and $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. Then, the following conditions are equivalent:*

- (i) $\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \text{OLSE}(\mathbf{X}\boldsymbol{\beta})$,
- (ii) $\mathcal{N}(\mathbf{H}) = \mathcal{N}(\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}) \cap \mathcal{N}(\mathbf{H})$.

Proof First we show that (ii) implies (i). The condition (ii) ensures that $\mathcal{N}(\mathbf{H}) \subseteq \mathcal{N}(\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}})$, which yields $\mathcal{R}(\overline{\mathbf{P}_{\mathbf{V}\mathbf{M}}}\mathbf{H}) \subseteq \mathcal{R}(\mathbf{H})$. Thus, $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H} = \overline{\mathbf{P}_{\mathbf{V}\mathbf{M}}}\mathbf{H}$. In consequence, $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H} = \mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H}$, and taking the transpose leads to $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H}$. The implication now follows on account of point (v) of Theorem 5.

The part (i) \Rightarrow (ii) is established in a similar fashion by reversing the preceding chain. \square

From the discussion preceding Corollary 1 it follows that $\mathcal{N}(\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}})$, specified in (10), coincides with $\mathcal{R}(\mathbf{P}_{\mathbf{V}\mathbf{M}})$ when (4) holds. In such a case, the condition (ii) of Theorem 6 reduces to $\mathcal{N}(\mathbf{H}) = \mathcal{R}(\mathbf{P}_{\mathbf{V}\mathbf{M}}) \cap \mathcal{N}(\mathbf{H})$, or, equivalently, to $\mathcal{R}(\mathbf{P}_{\mathbf{V}\mathbf{M}}) \subseteq \mathcal{N}(\mathbf{H})$, i.e., $\mathcal{R}(\mathbf{V}\mathbf{M}) \subseteq \mathcal{R}(\mathbf{M})$. When \mathbf{V} is nonsingular, we get $\mathcal{R}(\mathbf{V}\mathbf{M}) = \mathcal{R}(\mathbf{M})$, which is the final condition of Kruskal's Theorem in Krämer (1980).

Another observation is that the conditions of Theorems 5 and 6, unlike the customary conditions given on the top of the present section, predominantly deal with orthogonal projectors. Thus, the equality of $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and $\text{OLSE}(\mathbf{X}\boldsymbol{\beta})$ is characterized in a more symmetric way.

When $\mathbf{P}_{\mathbf{V}\mathbf{M}}$ has representation (6), then we get the following equivalences among the statements of Theorem 5:

- $\mathbf{H} + \mathbf{P}_{\mathbf{V}\mathbf{M}}$ is an orthogonal projector if and only if $\mathbf{A} = \mathbf{0}$,
- $\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{P}_{\mathbf{V}\mathbf{M}}\mathbf{H}$ if and only if $\mathbf{B} = \mathbf{0}$.

Note that condition $\mathbf{A} = \mathbf{0}$ in general is stronger than $\mathbf{B} = \mathbf{0}$, but in the present set-up they are equivalent. There exists a large number of equivalent conditions to characterize condition (v) of Theorem 5, for instance:

$$\begin{aligned} \mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}} &= \mathbf{P}_{\mathcal{R}(\mathbf{H}) \cap \mathcal{R}(\mathbf{V}\mathbf{M})}, \\ \mathcal{R}(\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}) &= \mathcal{R}(\mathbf{H}) \cap \mathcal{R}(\mathbf{V}\mathbf{M}), \\ \text{rank}(\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}) &= \dim[\mathcal{R}(\mathbf{H}) \cap \mathcal{R}(\mathbf{V}\mathbf{M})], \\ \text{rank}(\mathbf{H} - \mathbf{P}_{\mathbf{V}\mathbf{M}}) &= \text{rank}(\mathbf{H} + \mathbf{P}_{\mathbf{V}\mathbf{M}}) - \text{rank}(\mathbf{H}\mathbf{P}_{\mathbf{V}\mathbf{M}}); \end{aligned}$$

see Baksalary and Trenkler (2008).

Acknowledgments The authors are very grateful to two anonymous referees whose remarks considerably improved the paper. In particular, we are thankful for the hints concerning uniqueness of the BLUE from which, besides a number of improvement proposals, we got the idea for the example given in the Appendix.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix

As an example to describe the situation when $\mathcal{R}(\mathbf{X} : \mathbf{V})$ is a proper subset of $\mathbb{R}^{n,1}$, consider the linear model (1), where $\mathbf{X} = (1, 0, 0)' \in \mathbb{R}^{3,1}$ and $\mathbf{V} = \text{diag}(0, 1, 0) \in \mathbb{R}^{3 \times 3}$. Then $\mathcal{R}(\mathbf{X} : \mathbf{V})$ is the linear combination of the vectors $(1, 0, 0)'$ and $(0, 1, 0)'$, and does not fill out the whole space $\mathbb{R}^{3,1}$. It follows that $\mathbf{H} = \mathbf{X}\mathbf{X}^\dagger = \text{diag}(1, 0, 0)$, $\mathbf{M} = \mathbf{I}_3 - \mathbf{H} = \text{diag}(0, 1, 1)$, $\mathbf{V}\mathbf{M} = \mathbf{V}$, $\mathbf{H}\mathbf{V} = \mathbf{V}\mathbf{H} = \mathbf{0}$, $\mathbf{P}_{\mathbf{V}\mathbf{M}} = \mathbf{V}$, $\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}} = \text{diag}(1, 0, 1)$, $\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H} = \mathbf{H}$, and $\mathbf{T} = (\bar{\mathbf{P}}_{\mathbf{V}\mathbf{M}}\mathbf{H})^\dagger = \mathbf{H}$. Hence, $\mathbf{H}\mathbf{y} = \mathbf{T}\mathbf{y} = \mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^{3,1}$, i.e., $\text{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ everywhere.

Let us now have a look at the estimator $\mathbf{G}\mathbf{y}$, where $\mathbf{G} = \text{diag}(1, 0, g)$, with arbitrary $g \in \mathbb{R}$. The matrix \mathbf{G} satisfies Eq. (3), which means that with varying g the statistic $\mathbf{G}\mathbf{y}$ gives an infinite number of alternative representations of $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$. Observe that when $\mathbf{y} = (y_1, y_2, y_3)'$, then we get $(y_1, 0, gy_3)'$ as a best unbiased, but somewhat ridiculous estimator of $\mathbf{X}\boldsymbol{\beta}$. It follows that $\text{Cov}(\mathbf{H}\mathbf{y}) = \text{Cov}(\mathbf{T}\mathbf{y}) = \text{Cov}(\mathbf{G}\mathbf{y}) = \mathbf{0}$. Furthermore, when $g \neq 0$, then we have $\mathcal{N}(\mathbf{H} - \mathbf{G}) = \text{span}\{(1, 0, 0)', (0, 1, 0)'\}$, in contrast to $\mathcal{N}(\mathbf{H} - \mathbf{T}) = \mathbb{R}^{3,1}$. Note however that in this case the vector \mathbf{y} does not satisfy the consistency condition of Sect. 1.

References

- Baksalary JK, Kala R (1980) A new bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann Stat* 8:679–681
- Baksalary JK, Baksalary OM, Szulc T (2002) A property of orthogonal projectors. *Linear Algebra Appl* 354:35–39
- Baksalary OM, Trenkler G (2008) An alternative approach to characterize the commutativity of orthogonal projectors. *Discuss Math Probab Stat* 28:113–137
- Baksalary OM, Trenkler G (2009) A projector oriented approach to the best linear unbiased estimator. *Stat Pap* 50:721–733
- Baksalary OM, Trenkler G (2010) Functions of orthogonal projectors involving the Moore–Penrose inverse. *Comput Math Appl* 59:764–778
- Baksalary OM, Trenkler G (2011) Between OLSE and BLUE. *Aust N Z J Stat* 53:289–303
- Baksalary OM, Trenkler G (2012) On projectors and some of their applications in statistics. In: Bapat RB, Kirkland S, Prasad KM, Puntanen S (eds) *Lectures on matrix and graph methods*. Manipal University Press, Manipal, pp 113–127
- Baksalary OM, Trenkler G (2013) On column and null spaces of functions of a pair of oblique projectors. *Linear Multilinear Algebra*. doi:[10.1080/03081087.2012.731055](https://doi.org/10.1080/03081087.2012.731055)
- Greville TNE (1974) Solutions of the matrix equation $\mathbf{XAX} = \mathbf{X}$, and relations between oblique and orthogonal projectors. *SIAM J Appl Math* 26:828–832
- Groß J (2004) The general Gauss–Markov model with possibly singular dispersion matrix. *Stat Pap* 45:311–336
- Groß J, Trenkler G, Werner HJ (2001) The equality of linear transforms of the ordinary least squares estimator and the best linear unbiased estimator. *Sankhyā* 63:118–127
- Jaeger A, Krämer W (1998) A final twist on the equality of OLS and GLS. *Stat Pap* 39:321–324
- Krämer W (1980) A note on the equality of ordinary least squares and Gauss–Markov estimates in the general linear model. *Sankhyā* 42:130–131
- Krämer W, Bartels R, Fiebig DG (1996) Another twist on the equality of OLS and GLS. *Stat Pap* 37: 277–281

- Kruskal W (1968) When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Ann Math Stat* 39:70–75
- Puntanen S, Styan GPH, Isotalo J (2011) *Matrix tricks for linear statistical models: our personal top twenty*. Springer, Heidelberg
- Rao CR (1973) *Linear statistical inference and its applications*, 2nd edn. Wiley, New York
- Rao CR, Mitra SK (1971) *Generalized inverse of matrices and its applications*. Wiley, New York
- Trenkler G (1994) Characterizations of oblique and orthogonal projectors. In: Caliński T, Kala R (eds) *Proceedings of the international conference on linear statistical inference LINSTAT'93*. Kluwer, Dordrecht, pp 255–270